

✓

②

**AD-A242 432**



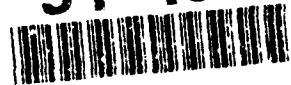
ALL INFORMATION CONTAINED  
HEREIN IS UNCLASSIFIED  
DATE 01-11-91 BY 1043  
C D

**Technical Report 1446**  
July 1991

# **Detection of Prosodics by Using a Speech Recognition System**

N. A. Hupp

**91-15174**



Approved for public release; distribution is unlimited.

# **NAVAL OCEAN SYSTEMS CENTER**

## **San Diego, California 92152-5000**

---

**J. D. FONTANA, CAPT, USN**  
**Commander**

**R. T. SHEARER, Acting**  
**Technical Director**

### **ADMINISTRATIVE INFORMATION**

The work reported here was performed during FY89 and 90 by the User Interface Technology Branch (Code 441), Command Support Technology Division, Command and Control Department, for the Office of Chief of Naval Research (OCNR-10P), Independent Research Programs (IR), Arlington, VA 22217-5000. As part of project ZW21, Detection of Prosodics in Speech Recognition Systems, the study was supported principally by the Independent Research Program at the Naval Ocean Systems Center.

Released by  
C. M. Dean, Head  
User Interface  
Technology Branch

Under authority of  
J. D. Grossman, Head  
Command Support  
Technology Division

### **ACKNOWLEDGMENTS**

The author would like to thank Suzanne Bemis, NOSC Code 441, who helped both with the experimental design and running the statistics for this research. Thanks also to Kevin Landel of the General Atomics San Diego Supercomputer Center at the University of California for his help in the linguistic aspects of the research and for supplying many references on prosodic research.

# **SUMMARY**

## **PROBLEM**

Determine the ability of a speech recognizer to extract prosodic speech features, such as pitch and stress, and to examine these features for application to future voice recognition systems.

## **RESULTS**

The Speech Systems Incorporated (SSI) speech recognizer demonstrated that it could detect prosodic features and that these features do indicate the word and/or syllable that is stressed by the speaker. The research examined the effect of prosodics, such as pitch, amplitude, and duration, on word and syllable stress by using the SSI. Subjects read phrases and sentences, using a given intonation and stress. The three sections of the experiment compared questions and answers, words stressed within a sentence, and noun/verb pairs, such as "object" and "subject." The results were analyzed both on the syllable level and the word level. In all cases, there was a significant increase in pitch, amplitude, and duration when comparing stressed words and syllables to unstressed words and syllables. When comparing unstressed words only, it was also noted that the first word in a sentence has an increase in pitch, amplitude, and duration. The threshold could be set in recognition systems for each of these parameters. Current speech recognizers do not use acoustic data above the word level. This research shows that we have the capability of developing better speech systems by incorporating prosodics with new linguistic software.

## **RECOMMENDATIONS**

Further research should be directed to answering the following questions:

- Which speech recognizers besides the SSI can detect prosodic differences?
- How does word placement in a sentence affect stress?
- Do particular phonemes or phoneme combinations in the sentence affect stress?
- Does the second syllable of a noun/verb have shorter relative duration if the word is in the first, middle, last, or only position of a sentence?

## CONTENTS

|  |    |
|--|----|
| INTRODUCTION .....   | 1  |
| PURPOSE .....  | 1  |
| BACKGROUND .....   | 1  |
| METHOD .....   | 4  |
| SUBJECTS .....   | 4  |
| EQUIPMENT .....  | 4  |
| EXPERIMENTAL DESIGN .....  | 4  |
| DATA COLLECTION .....  | 5  |
| RESULTS .....  | 5  |
| QUESTIONS AND STATEMENTS .....   | 5  |
| STRESSED/UNSTRESSED WORDS .....  | 6  |
| NOUN/VERB PAIRS .....  | 10 |
| DISCUSSION .....   | 13 |
| RECOMMENDATIONS .....  | 15 |
| CONCLUSION .....   | 15 |
| REFERENCES .....   | 16 |
| BIBLIOGRAPHY .....   | 17 |
| APPENDIX .....   | 18 |
| <b>FIGURES</b>   |    |
| 1. Using the phonemes from That. and That? the pitch and amplitude were graphed against their duration .....           | 7  |
| 2. The pitch and amplitude were graphed against their duration, using questions and statements on the word level ..... | 8  |
| 3. The pitch, amplitude, and duration are depicted here, using questions and statements on the word level .....        | 9  |
| 4. The pitch and amplitude were graphed against their duration, using statements containing stressed words .....       | 11 |

## CONTENTS (continued)

|   |    |
|---|----|
| 5. The pitch, amplitude, and duration are depicted here, using statements containing stressed words .....             | 12 |
| 6. Duration of syllables (in ms) .....  | 13 |
| 7. Using the phonemes from sub-ject' and sub'-ject, the pitch and amplitude were graphed against their duration ..... | 14 |



|                    |                                     |
|--------------------|-------------------------------------|
| Accession For      |                                     |
| NTIS Grant         | <input checked="" type="checkbox"/> |
| DTIC Tab           | <input type="checkbox"/>            |
| ADONIS             | <input type="checkbox"/>            |
| Justification      | <input type="checkbox"/>            |
| By                 |                                     |
| Distribution       |                                     |
| Availability Codes |                                     |
| Dist               | Avail and/or Special                |
| A-1                |                                     |

## INTRODUCTION

When one recognizes speech, one does not just recognize words, but the words together with their stress, intonation, pauses, and timing in speech. "**Prosody**" is concerned with these additional features in speech. **Stress** is the most basic and important feature. **Intonation** refers to the pitch of the speech, and this also provides information to aid understanding. Some examples in which the use of prosody helps to avoid ambiguity include:

What is in the road ahead?

What is in the road? A head? (Differences in **intonation**.)

We fed (her dog) biscuits.

We fed her (dog biscuits). (Differences in **stress** and pauses.)

Lighthouse keeper/light housekeeper. (Differences in **stress**.)

Detecting prosodics provides additional meaning to a sentence (as shown in the examples). Prosodic information clears up ambiguities, can reflect the emotions and attitudes of the speaker, and aids in understanding of natural language.

Although research has been conducted on how humans use prosodics in understanding spoken language, only limited research has been conducted in using speech recognition systems to detect prosodics. This project analyzed the effect of prosodics on word and syllable stress by using a Speech Systems Incorporated (SSI) recognizer.

## PURPOSE

The objective of this project was to determine the extractability of prosodic features, such as pitch and stress, through a speech recognizer, and to examine these features to determine whether a voice recognition system can accurately detect prosodics. The ultimate goal was to establish whether these features could indicate that a word or phoneme was stressed. If so, this information could be used in the development of better recognition systems, in which prosodic features could help indicate the meaning (semantics) of a phrase. In a structured syntax, for example, it could be possible for a voice recognition system to use these prosodic features to determine which syntactic path best suits the stated phrases.

## BACKGROUND

A literature search, as well as discussions with personnel from the Defense Advanced Research Projects Agency (DARPA), indicated that limited direct research

had been performed on using commercial off-the-shelf speech recognition systems to detect prosodics. Based on these discussions, the literature review, and the importance of prosodics to speech understanding, the current research program was initiated.

Other researchers have recognized the importance of studying prosodics. Wayne Lea (1980) points out the importance of studying prosodic features:

"If there is one aspect of the information in the speech signal that seems promising and yet 'untapped,' it is the 'prosodic' information such as stress patterns, intonation, pauses, and timing structures in the speech. Here a newcomer to speech recognition studies can readily make an original contribution, and experienced speech recognition studies can find additional tools for substantially improving the performance of speech understanding systems. Prosodic analysis is one of the 'gaps' in speech recognition technology that has been repeatedly (and increasingly) noted since work on sentence understanding systems began."

Both Lea and Waibel have come up with strategies for detecting prosodics. An algorithm for locating stressed syllables from fundamental frequency contours and high-energy syllabic nuclei correctly located the nuclei of over 85 percent of all those syllables perceived as stressed by a panel of listeners (Lea, Medress, and Skinner, 1975). Major attempts at using prosodic cues in speech recognition have focused on syntactic analysis. These attempts have aided syntactic analysis by setting phrase or clause boundaries through the prosodics of pitch and stress (Waibel 1987).

A listener's knowledge of the constraints in the English language enables one to infer where one word ends and the next begins. Listeners know how the stress pattern and rhythm in English sentences constrain the possible parsing of a sentence into words. This knowledge is used "to hear" where the words begin and end.

Nahatani and Schaffer (1978) ran a series of four experiments. In the first experiment, they substituted the syllable "ma" for adjective-noun phrases in short sentences. The stress was either placed on the first "ma" or the second "ma" in "mama." Then phrases were recorded for seven talkers. Ten subjects listened to nine phrases from each talker. Results indicated that it was possible for subjects to choose the syllable with the dominant stress.

In the second experiment, listeners judged whether they heard each phrase as "ma mama" or "mama ma". These authors found that the stress pattern was not a sufficient cue for parsing ambiguous phrases, yet the listeners parsed even these phrases more than 50% of the time. The authors concluded that there were other prosodic features, such as rhythm, glottal stops, and aspirations, that differentiated these ambiguous phrases and enabled listeners to parse them correctly.

In the third experiment, hybrid speech synthesis was used to study the prosodic features of stress, rhythm, pitch, amplitude, and spectrum as cues for parsing the

ambiguous phrases. Hybrid speech synthesis is a general experimental technique for assessing the strength and interactions with a set of speech features that influence speech perception. Hybrid speech synthesis was done by means of linear predictor analysis and synthesis, so that the prosodic features of speech could be accessed and manipulated independently. The linear predictor parameters consisted of pitch and amplitude parameters as well as 12 parameters representing the spectral feature. The rhythm feature was represented by frames at 10-ms intervals that specified values for the other 14 parameters. The parameters representing the natural features of the parent phrases were obtained directly from natural speech by linear predictor analysis of the ambiguous phrases of the first parsing experiment. The results showed that the parsing of a phrase was affected when its rhythmic pattern was changed, but not when its pitch and amplitude contours were changed.

In the final experiment, the durations of the two phonemes were normalized so that every talker had the same speaking rate. This indicated that speech rhythm correlated with the stress pattern and parsing of phrases. Nahatani and Schaffer concluded that the stress pattern and speech rhythm are primary prosodic cues for word perception.

Lehiste (1970) found that the greater the degree of stress on a word in a sentence, the greater the duration of the vowel of the word. Furthermore, the pitch and the amplitude of the stressed syllable is higher. In addition, Cutler (1976) found that words with little stress in a sentence tend to have shorter duration, lower pitch, and less amplitude.

Landel (1983) performed several experiments in prosodics, using a speech recognizer. In the first part of his research, he demonstrated that people can assimilate prosodic features to extract semantics from a linguistic phrase. Using one-syllable words (for example, 'there' and 'that'), he showed people could hear the difference between questions and statements. With a Texas Instruments (TI) speech processing card in an IBM PC, he showed that variations in pitch could be detected by a machine system. The sentences and phrases used in his and other research were selected for the research described in this report, based on the structure of the sentences.

These studies suggest that continued research on prosodics could improve future speech systems. Accuracy could be improved for voice input to computers, and synthesized speech could sound more natural.

Based on the literature review, it is hypothesized that the SSI will not only be able accurately to detect the pitch, amplitude, and duration of words, but also to indicate differences in pitch, amplitude, and duration between stressed and unstressed words.

## METHOD

### SUBJECTS

Ten subjects were used for the experiment, three female speakers and seven male speakers. All were scientists, ages 23 through 54.

### EQUIPMENT

Just as letters are the basic units of written text, phonemes are the basic units of spoken speech. For example, the words "bit" and "pit" are distinguished from each other by the phonemes \b\ and \p\, respectively. Today, most recognizers work at the word level. The Speech Systems Incorporated recognition system was chosen for this project because it is a phonetic recognition system that recognizes each individual phoneme and is therefore capable of recognizing continuous speech sounds beyond single words. The SSI includes C-language functions in the phonetic decoder interface (PDI). The PDI functions include software to extract the pitch, duration, and amplitude of each phoneme of a given sentence.

### EXPERIMENTAL DESIGN

Each subject read single-word phrases or five-word sentences, which were presented on a Sun Workstation. Instructions were also included on how each sentence should be spoken.

An example of what a subject would see on the computer screen might include:

*Please say : Mike opened THAT green door.*

*As though you are answering the question: Mike opened which green door?*

All of the sentences had been used in previous research in prosodics. The words and phrases were especially chosen from the literature as easily segmented and evaluated. Two kinds of sentences were used: one-word sentences and five-word sentences. Using one-word sentences made it easier to isolate a word for evaluation. The longer, five-word sentences were chosen so that every word ended in a stop, (for example, 'k', 't'). Therefore, words could not be slurred together, and segmentation during evaluation would be less difficult.

The first phase of the experiment concentrated on statement and question pairs. For example:

There.

There?

Jack cooked that big fish.

(Statement, no stress on any words.)

Jack cooked that big fish?

(Question, no stress on any words.)

The second phase concentrated on stressed versus unstressed words within a sentence. For example:

JACK cooked that big fish.

(Who cooked that big fish?)

Mike opened that GREEN door.

(Mike opened what color door?)

The final phase concentrated on noun/verb pairs, which are distinguished from each other almost entirely on the basis of stress (Lea, 1978).

PERmit/PerMIT

OBject/ObJECT

SUBject/SubJECT

For a listing of all the sentences used in the experiment, please see the appendix.

Each subject was tape recorded to keep a record of how each sentence was spoken.

## DATA COLLECTION

The segments into which the SSI divides the speech input are called transems. A transeme is SSI's representation of fragments of a phoneme. The output from the experiment showed the spoken speech broken down into transems. There was a pitch, amplitude, and duration associated with each transeme. The actual data output from the SSI cannot be included in this report, due to the proprietary nature of the software.

## RESULTS

The BMDP statistical package was used to determine whether there was any significance in (1) the various parameters of the question and answer pairs; (2) the stressed versus the unstressed words of the five-word sentences; and (3) the syllables of the noun/verb pairs.

## QUESTIONS AND STATEMENTS

The results from question/statement pairs gave prosodic data at the transeme level. The amplitudes are greatest during the vowel sounds. Durations in the example, and in the other question/statement pairs, were all approximately equal.

Analysis showed that the pitch was significantly higher at the end of a question than at the end of a statement. Duration was significantly longer on the word level. The word "there" was longer than the word "that." Duration was longer on the phoneme level. Vowels had longer durations than consonants. Amplitude was also significantly higher on the phoneme level for vowels. These data are shown in figures 1, 2, and 3.

Figure 1 shows "that?" as a question, and "that" as a statement. Each phoneme had an amplitude and a pitch, which was graphed against its duration. Both pitch and amplitude tended to increase at the end of the word when the question was uttered and decrease when a statement was spoken.

Figure 2 depicts sentences on the word level. Again, the pitch and amplitude showed a greater increase in the question than the statement. The last word of the sentence in the statement had a longer duration.

Figure 3 shows the same data in a bar-graph form. This more clearly demonstrates that the duration increases in the statement form.

### **STRESSED/UNSTRESSED WORDS**

Evaluating the stressed/unstressed parameter of the experiment involved careful definition of the dependent variable. After selecting the appropriate transems and the parameters to go with them, the transems were separated into voiced and unvoiced sounds to obtain meaningful pitch data. Voiced sounds have a pitch, whereas the pitch on an unvoiced sound is meaningless. Thus only the voiced sounds were used for pitch. The mean pitch of all voiced sounds in each word was then analyzed. All transems were used for the amplitude and duration. In the five-word sentences, the duration for the transems in each word were summed together and the amplitude of all transems in each word were averaged. Thus the dependent measures were analyzed on the word level. The BMDP statistical package then was used to perform a two-factor (stress  $\times$  word), within-subject analysis of variance (ANOVA) on each of the dependent variables (pitch, amplitude, and duration).

The results showed that pitch ( $F = 8.08$ ,  $P = 0.02$ ), amplitude ( $F = 20.55$ ,  $P < 0.01$ ), and duration ( $F = 35.34$ ,  $P < 0.01$ ) significantly increased when a word was stressed, as compared to the same word unstressed. The increase was 9% in pitch, 43% in amplitude, and 55% in duration.

There was a stress level by word ( $S \times W$ ) interaction for pitch ( $F = 2.83$ ,  $P = 0.01$ ) and for amplitude ( $F = 2.75$ ,  $P = 0.01$ ). This interaction showed that pitch was higher for the first word (Jack, Mike) of a sentence, as compared to the last word of the sentence (fish, door).

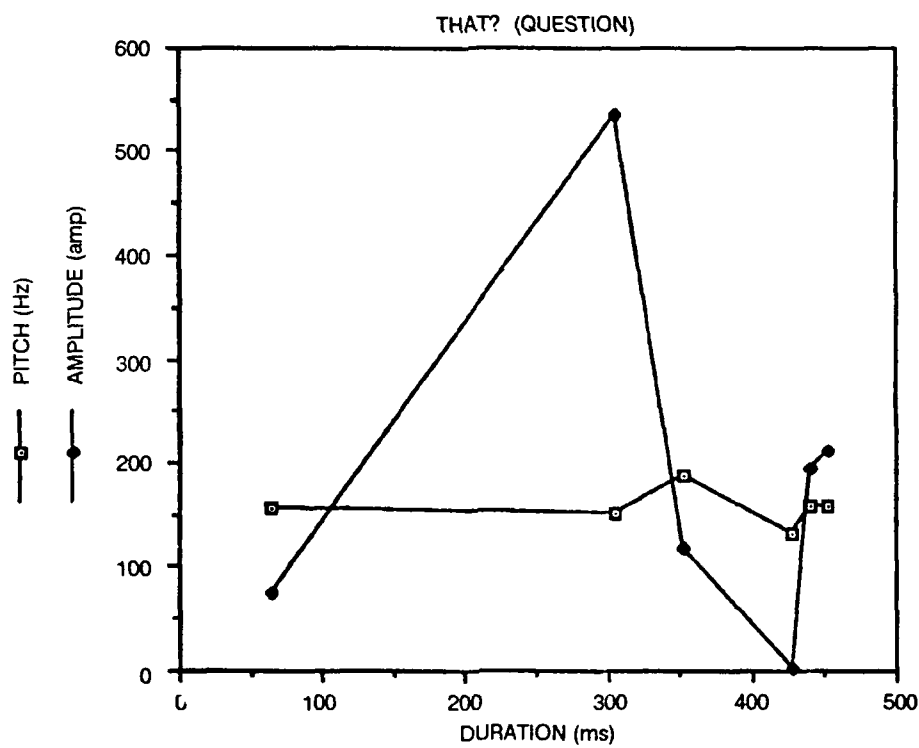
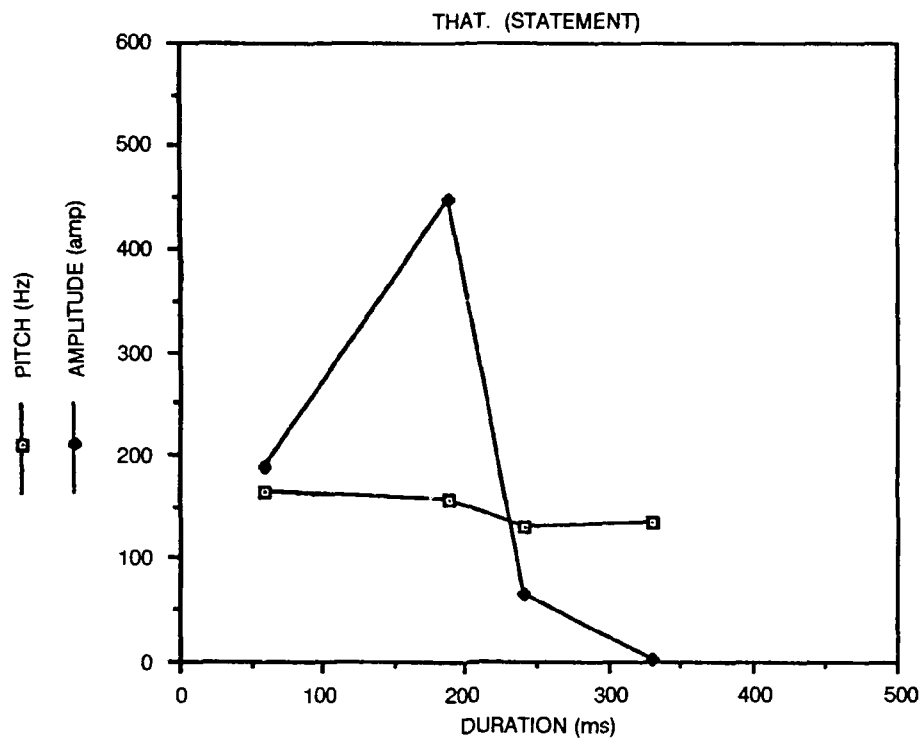


Figure 1. Using the phonemes from That. and That? the pitch and amplitude were graphed against their duration.

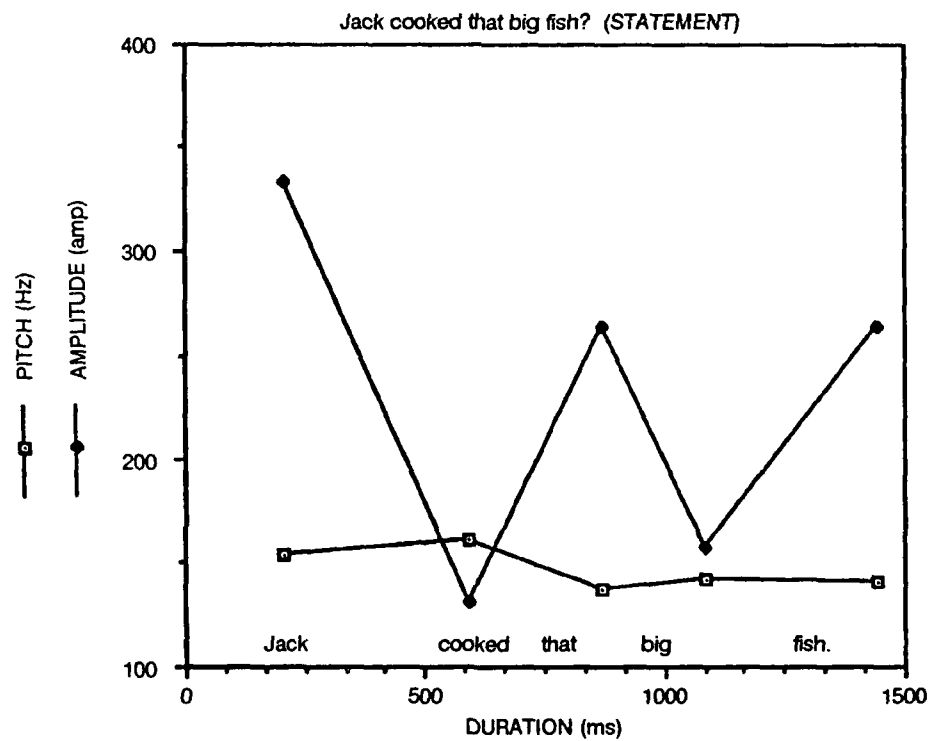
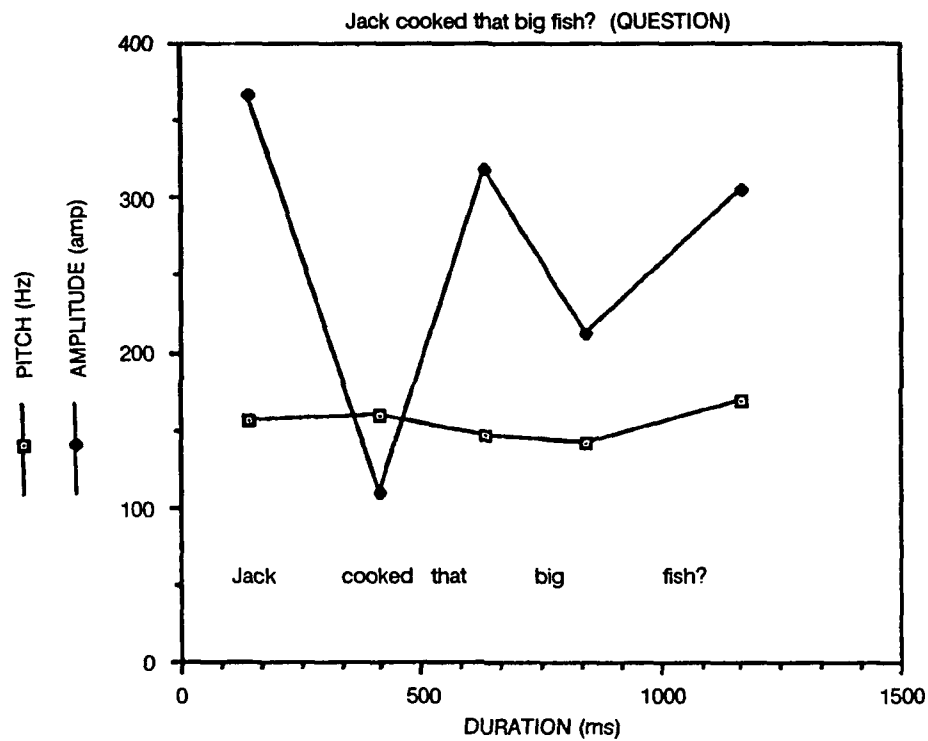


Figure 2. The pitch and amplitude were graphed against their duration, using questions and statements on the word level.

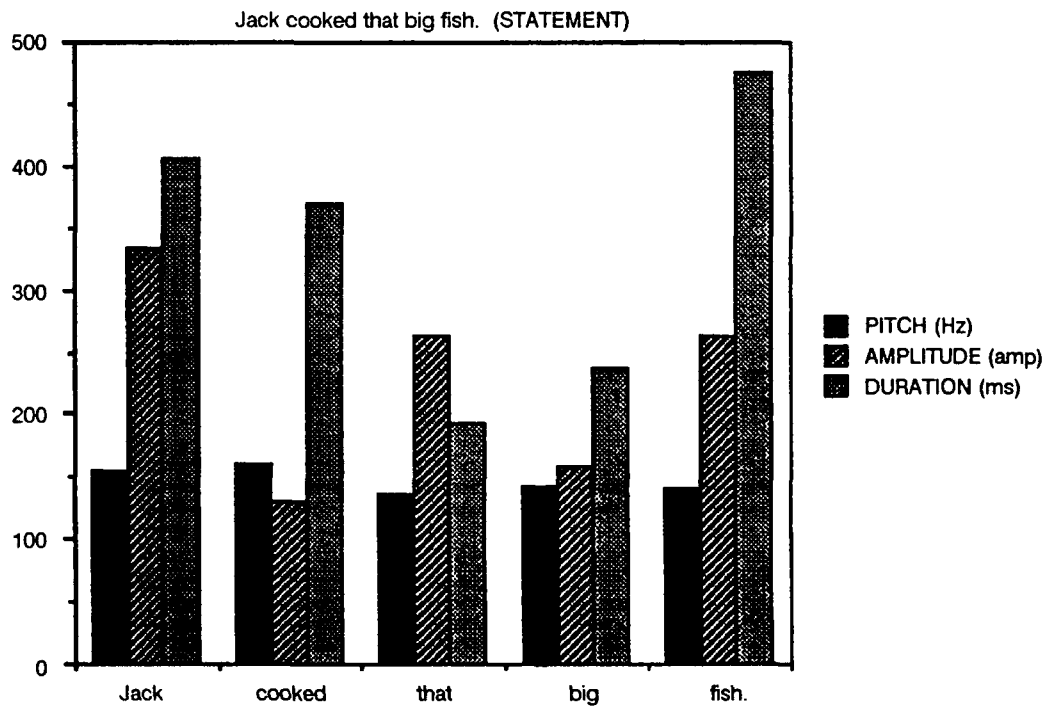
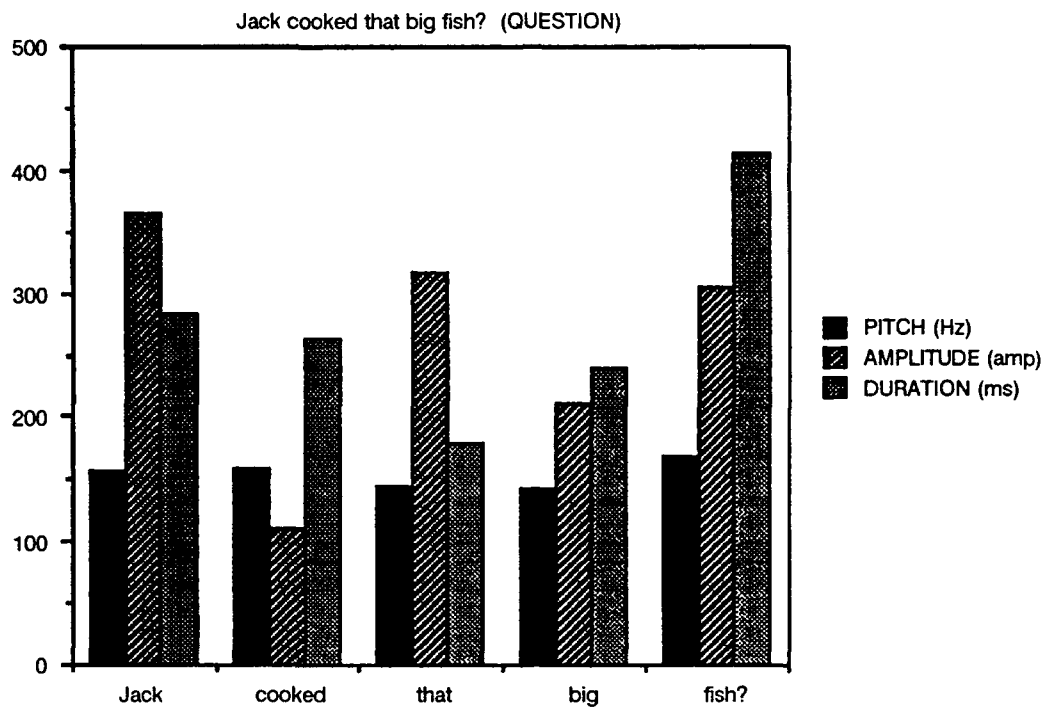


Figure 3. The pitch, amplitude, and duration are depicted here, using questions and statements on the word level.

A regression and correlational analysis showed that there was no correlation between pitch, amplitude, and duration. The increase in pitch did not correlate with the increase in amplitude or duration. Although all stressed words had a higher pitch and amplitude, as well as a longer duration, than unstressed words, the relative rise in pitch, increase in amplitude, and length of duration were all independent of each other.

Figures 4 and 5 show two examples of stressed words and their impact on pitch, amplitude, and duration. Data are shown in both line-graph, and bar-graph form.

### **NOUN/VERB PAIRS**

Syllables were analyzed for noun/verb pairs. A two-factor analysis of variance was performed on the syllables in the words "permit," "object," and "subject." The two factors were (a) stressed or unstressed condition and (b) syllables. Results of the analysis showed that pitch and amplitude were significantly higher and duration significantly longer for the stressed syllables. These findings are consistent with the stressed words in the analysis above.

An unexpected finding was noted in this study. The duration for all six syllables was compared in the stressed condition to determine whether any significant duration differences existed between the first and second syllable. In other words, the durations for the stressed OB of "object," the stressed SUB of "subject," and the stressed PER of "permit" were compared to the condition when the second syllables (JECT and MIT) were stressed. Contrary to expectation, the duration of the unstressed second syllable was always greater than the stressed first syllable. The results of the comparison of all stressed syllables showed that the second syllable, when stressed, always had a greater duration than when the first syllable was stressed. See figure 6.

Although the pitch always increased on the stressed syllable (comparing stressed and unstressed syllables), the pitch decreased on the second syllable when the comparison was made between a stressed first syllable and a stressed second syllable. The amplitude was only higher when the syllable was stressed. Figure 7 shows subject as a noun and as a verb. The relative duration increased in the stressed syllable as well as the relative pitch.

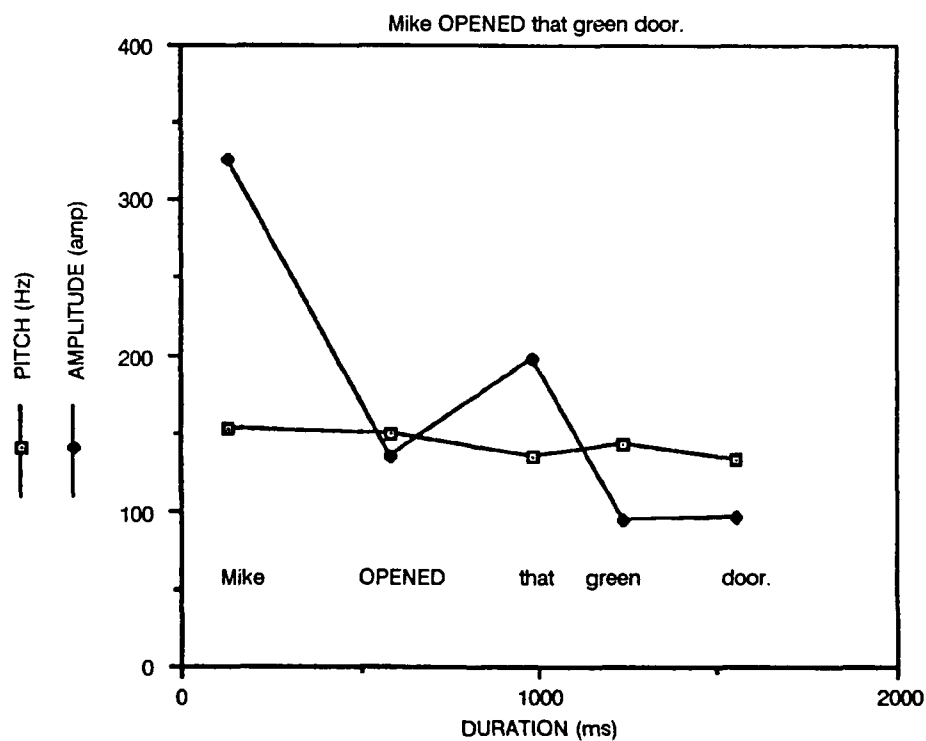
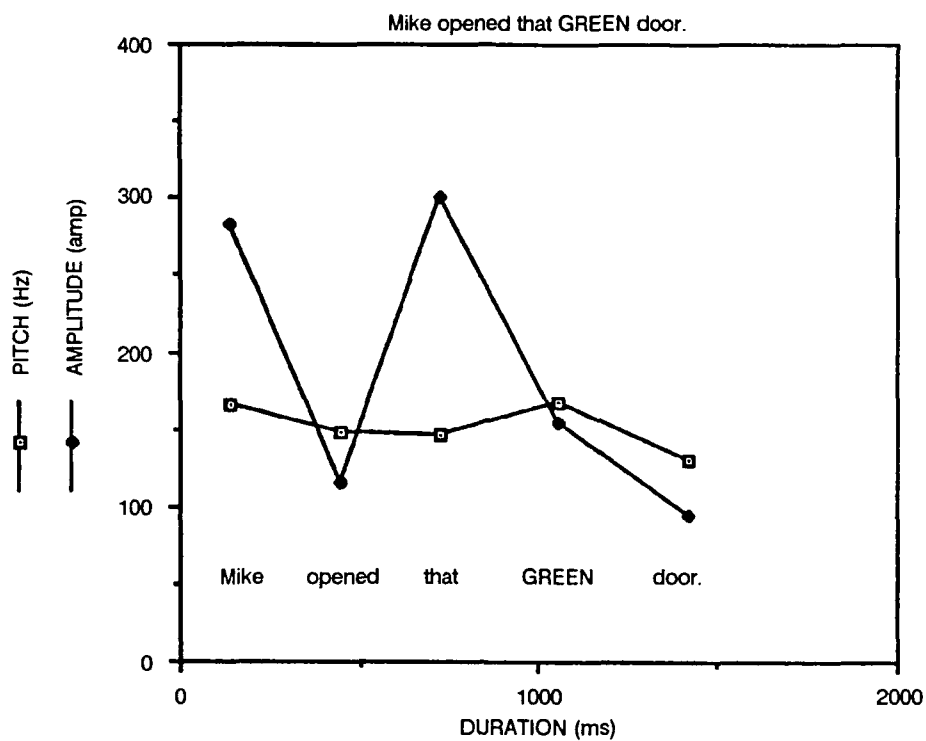


Figure 4. The pitch and amplitude were graphed against their duration, using statements containing stressed words.

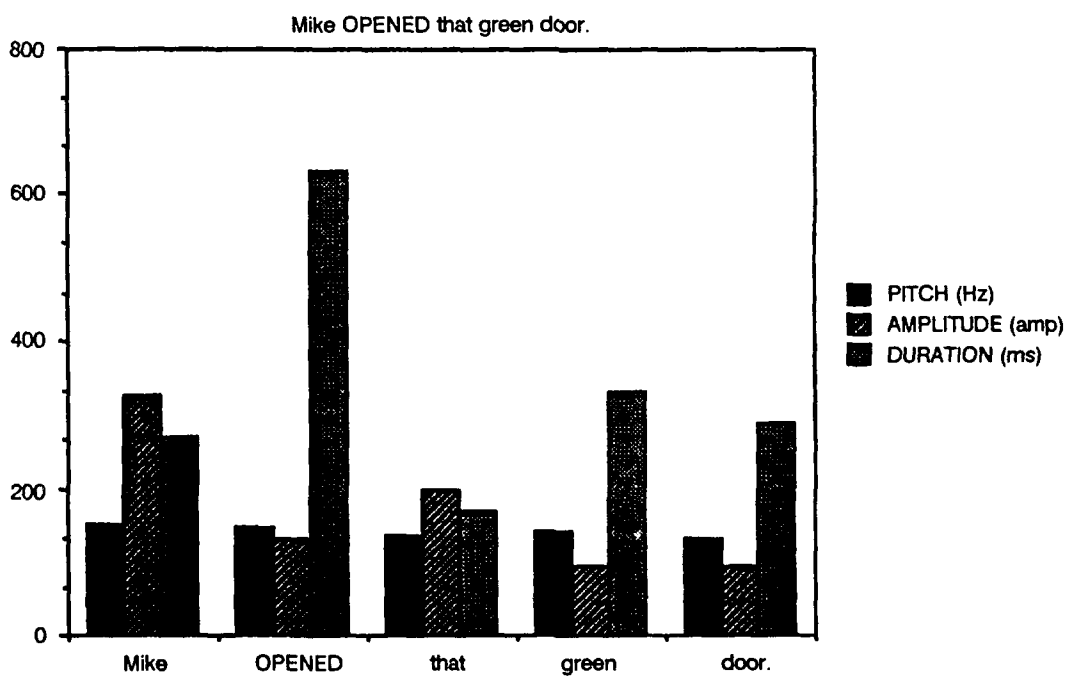
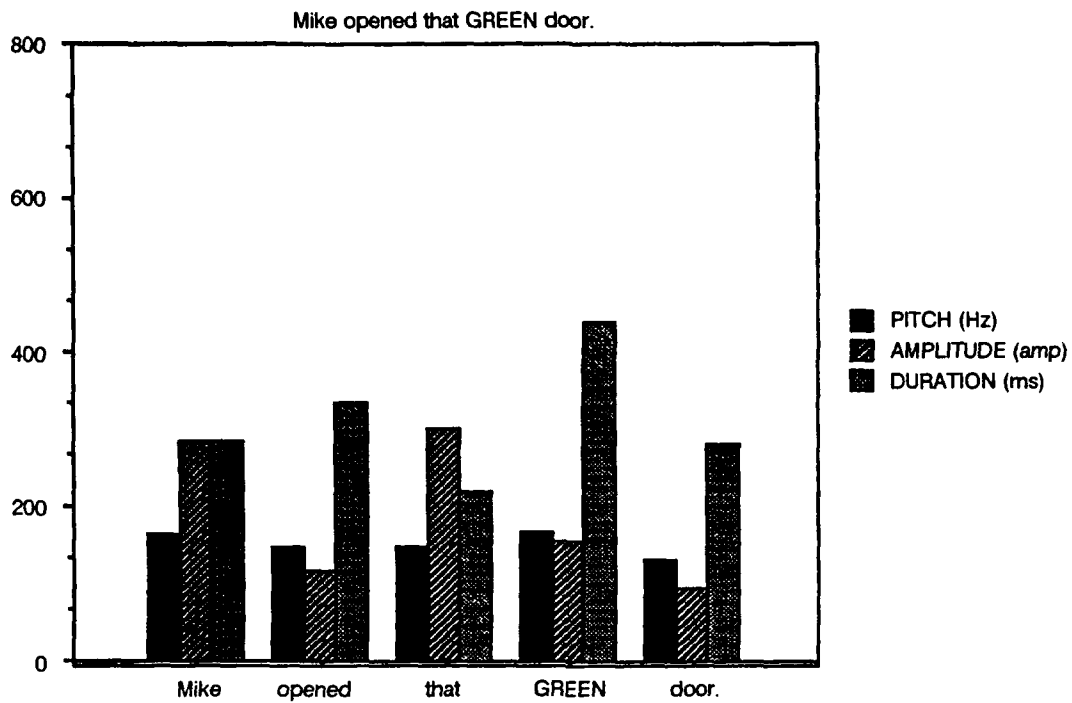


Figure 5. The pitch, amplitude, and duration are depicted here, using statements containing stressed words.

| <u>Stressed<br/>syllables</u> |             | <u>Unstressed<br/>syllables</u> |             |
|-------------------------------|-------------|---------------------------------|-------------|
| OB<br>214                     | JECT<br>459 | ob<br>155                       | ject<br>338 |
| SUB<br>290                    | JECT<br>478 | sub<br>245                      | ject<br>316 |
| PER<br>190                    | MIT<br>384  | per<br>198                      | mit<br>335  |

Figure 6. Duration of syllables (in ms).

## DISCUSSION

The hypothesis was found to be true; i.e., the Speech Systems Incorporated speech recognizer demonstrated that it can detect prosodic features and that these features do indicate the word and/or syllable that is stressed by the speaker. These results also confirm Landel's findings; i.e., speech systems are capable of detecting prosodics and of using these prosodics to indicate which word or syllable is stressed.

Findings by O'Shaughnessy (1987) showed that the last stressed syllable in a phrase usually has a longer duration and may account for the second syllable of the noun/verb pairs (subject, object, permit) always having a longer duration.

The last syllable, "mit," as in "permit," did not vary as much as "ject." This is attributed to the fact that many of the subjects pronounced the noun "permit" in the same way as the verb "permit."

There were some additional findings from the analyses. The stress by word interaction for amplitude showed that amplitude was consistently higher for the first word of a sentence (Jack, Mike), as compared to all other words within the sentence across stressed and unstressed conditions.

Curiously, in comparing all words in the unstressed condition, it was found that the middle word, "that," in both sentences had the second highest amplitude after "Jack" and "Mike." Since the speakers were not asked to emphasize this word, it is theorized that the style of the sentence suggested the emphasis.

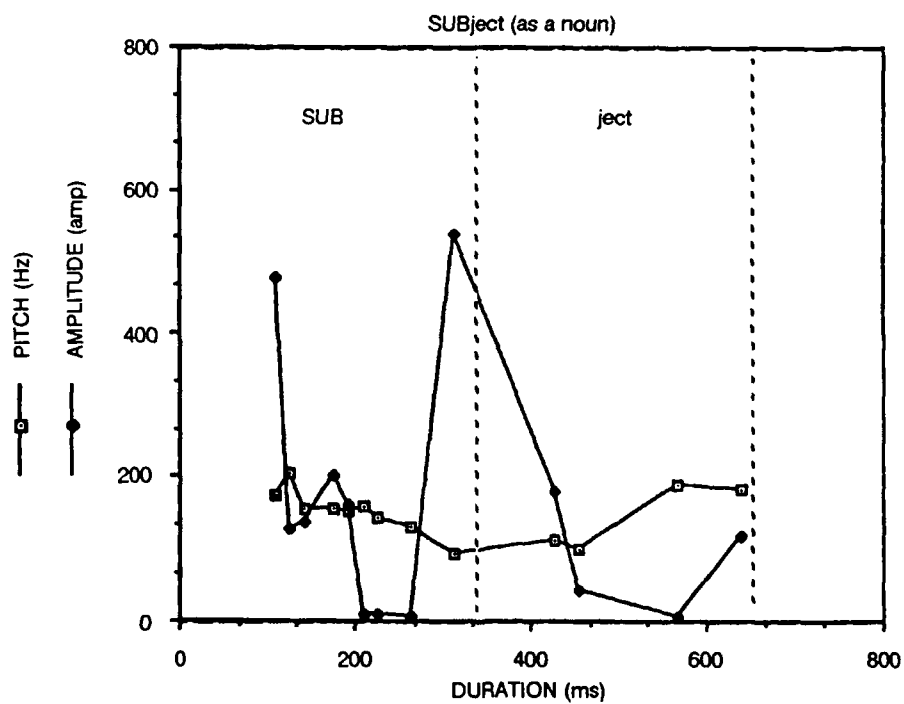
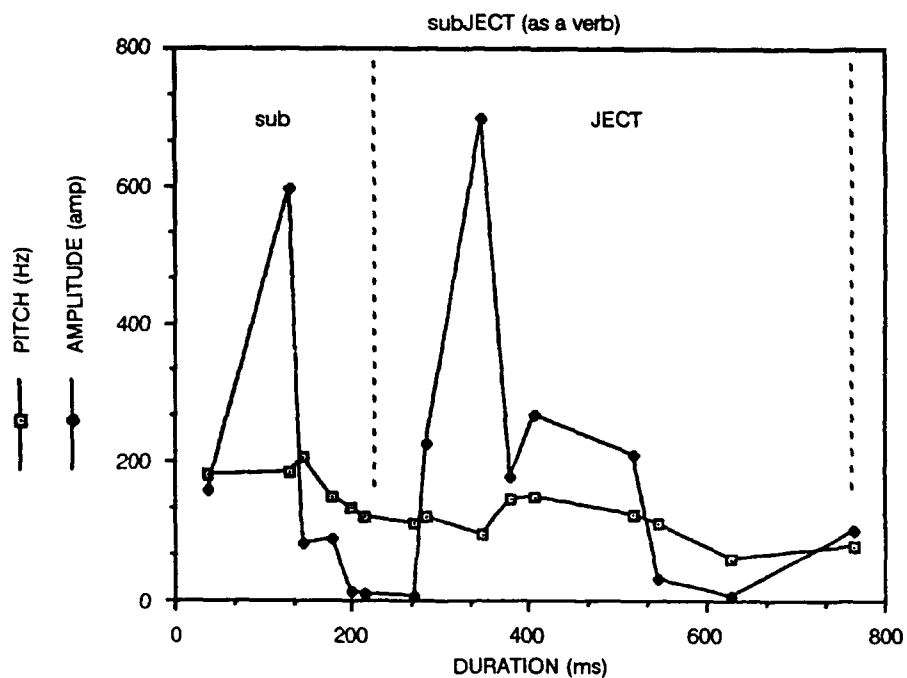


Figure 7. Using the phonemes from sub-ject' and sub'-ject, the pitch and amplitude were graphed against their duration.

Predictably, the duration was longer in the two-syllable words, "cooked" and "opened," than any of the other unstressed words. This makes sense because the word length is increased when more syllables are present in a word.

The significant increase in pitch, amplitude, and duration show that a threshold could be set in recognition systems for each of these prosodics. For example, the pitch, amplitude, and duration could be obtained for each word in a vocabulary for both stressed and unstressed words. Within a certain range, the recognizer would know whether the word was stressed and an expert system could then compare all words in the phrase so that the correct syntactic path is followed. Thus semantic understanding by the system through prosodics would improve recognition accuracy for any speech recognition system.

## **RECOMMENDATIONS**

Further research can help answer the following questions:

- Which speech recognizers besides the SSI can detect prosodic differences?
- How does word placement in a sentence affect stress?
- Do particular phonemes or phoneme combinations in the sentence affect stress?
- Does the second syllable of a noun/verb word have a shorter relative duration if the word is in the first, middle, last, or only position of a sentence?

## **CONCLUSION**

Although speech recognizers do not currently use acoustic data above the word level, this research shows that they have the capability of detecting important speech prosodics, including stress, intonation, and pauses. This capability, combined with advanced linguistic software, could significantly improve the performance of speech recognition systems.

## REFERENCES

- Cutler, A. 1976. "Phoneme-monitoring Reaction Time as a Function of Preceding Intonation Contour," *Perception and Psychophysics*, vol. 20, pp. 55-60.
- Landel, K. 1988. *On the Perception of Pitch Accents*, master's thesis, Massachusetts Institute of Technology, Cambridge, MA (September).
- Lea, W. 1980. *Trends in Speech Recognition*, Prentice-Hall, Inc., Englewood Cliffs, NJ, ch. 8, pp. 166-205.
- Lea, W., M. Medress, T. Skinner. 1975. "A Prosodically Guided Speech Understanding Strategy," *IEEE Trans. ASSP*, vol. ASSP-23, no. 1.
- Lehiste, I. 1970. *Suprasegmentals*, MIT Press, Cambridge, MA.
- Nahatani, L., J. Schaffer. 1978. "Hearing 'Words' Without Words: Prosodic Cues for Word," *J. Acoust. Soc. Am.* (January).
- Waibel, A. 1987. "Prosodic Knowledge Sources for Word Hypothesization in a Continuous Speech Recognition System," *Proc. IEEE ICASSP 87*, vol. 2, pp. 856-859.

## BIBLIOGRAPHY

Beach, C. 1989. *Duration and Pitch Combine to Represent Grammatical Structure in Temporarily Ambiguous Spoken Sentences*, doctoral dissertation, Univ. of Wisconsin, Madison, WI.

Klatt, D. 1976. "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence," *J. Acoust. Soc. Am.*, vol. 59, no. 5, p. 1208 (May).

Ladefoged, P. 1971. *Elements of Acoustic Phonetics*, Univ. of Chicago Press, Chicago, IL.

Landel, K. 1987. *Sing It Like This*, annual report on research sponsored by Nippon Telegraph and Telephone Company. MIT Press, Cambridge, MA.

Lea, W., and F. Clermont. 1984. "Algorithms for Acoustic Prosodic Analysis," *IEEE Trans. ASSP*, no. 1, p. 42.7.

Lehiste, I. 1967. *Readings in Acoustic Phonetics*, MIT Press, Cambridge, MA.

Nasre, M., G. Caelen-Haumont, and J. Caelen. 1989. "Using Prosodic Rules in Speech Recognition Expert System," *IEEE Trans. ASSP*, p. 671-674.

O'Shaughnessy, D. 1987. *Speech Communication - Human and Machine*, Addison-Wesley Publishing Co., Menlo Park, CA.

Pierrehumbert, J. 1989. *Prosody*, DARPA Speech and Natural Language Workshop, Philadelphia, PA (February).

Price, P., M. Ostendorf, C. Wightman. 1989. *Prosody and Parsing*, DARPA Speech and Natural Language Workshop, Cape Cod, MA (October).

Speech Systems Incorporated. 1987. *Speech Input Development System*, SSI Reference Manual 3.1, Tarzana, CA.

## APPENDIX

Here is a list of all the sentences and phrases used in the experiment.

The first phase of the experiment concentrated on statement and question pairs.

There.

There?

That.

That?

Jack cooked that big fish.

(Statement, no stress on any words.)

Jack cooked that big fish?

(Question, no stress on any words.)

Mike opened that green door.

(Statement, no stress on any words.)

Mike opened that green door?

(Question, no stress on any words.)

The second phase concentrated on stressed versus unstressed words within a sentence.

JACK cooked that big fish.

(Who cooked that big fish?)

Jack COOKED that big fish.

(What did Jack do with that big fish?)

Jack cooked THAT big fish.

(Which big fish did Jack cook?)

Jack cooked that BIG fish.

(Which fish did Jack cook?)

Jack cooked that big FISH.

(Did Jack cook that big fish or that big steak?)

MIKE opened that green door.

(Who opened that green door?)

Mike OPENED that green door.

(What did Mike do to that green door?)

Mike opened THAT green door.

(Mike opened which green door?)

Mike opened that GREEN door.

(Mike opened what color door?)

Mike opened that green DOOR.

(Mike opened the green what?)

The final phase concentrated on noun/verb pairs.

PERmit/PerMIT

OBject/ObJECT

SUBject/SubJECT.

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

|  |   |  |  |   |  |
|--|---|--|--|---|--|
| 1. AGENCY USE ONLY (Leave blank)   |   | 2. REPORT DATE<br>July 1991                                |  | 3. REPORT TYPE AND DATES COVERED<br>Final: Oct 89 — Sep 90  |  |
| 4. TITLE AND SUBTITLE<br><b>DETECTION OF PROSODICS BY USING A SPEECH RECOGNITION SYSTEM</b>  |   |  |  | 5. FUNDING NUMBERS<br>Proj: ZW21<br>WU: DN300019            |  |
| 6. AUTHOR(S)<br>N. A. Hupp   |   |  |  |   |  |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>Naval Ocean Systems Center<br>San Diego, CA 92152-5000   |   |  |  | 8. PERFORMING ORGANIZATION<br>REPORT NUMBER<br>NOSC TR 1446 |  |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>Office of Chief of Naval Research (OCNR-10P)<br>Independent Research Programs (IR)<br>Arlington, VA 22217-5000  |   |  |  | 10. SPONSORING/MONITORING<br>AGENCY REPORT NUMBER           |  |
| 11. SUPPLEMENTARY NOTES  |   |  |  |   |  |
| 12a. DISTRIBUTION/AVAILABILITY STATEMENT<br><br>Approved for public release; distribution is unlimited.  |   |  |  | 12b. DISTRIBUTION CODE                                      |  |
| 13. ABSTRACT (Maximum 200 words)<br><br>The ability of a speech recognizer to extract prosodic speech features, such as pitch and stress, and to examine these features for application to future voice recognition systems, was determined. The Speech Systems Incorporated (SSI) speech recognizer demonstrated that it could detect prosodic features and that these features do indicate the word and/or syllable that is stressed by the speaker. The research demonstrates the ability to develop better speech systems by incorporating prosodics with new linguistic software. |   |  |  |   |  |
| 14. SUBJECT TERMS<br>linguistics<br>prosodics<br>prosody speech recognizers<br>Speech Systems Incorporated (SSI)<br>voice recognition systems<br>pitch<br>stress   |   |  |  | 15. NUMBER OF PAGES<br>28                                   |  |
|  |   |  |  | 16. PRICE CODE  |  |
| 17. SECURITY CLASSIFICATION<br>OF REPORT<br>UNCLASSIFIED   | 18. SECURITY CLASSIFICATION<br>OF THIS PAGE<br>UNCLASSIFIED | 19. SECURITY CLASSIFICATION<br>OF ABSTRACT<br>UNCLASSIFIED | 20. LIMITATION OF ABSTRACT<br>SAME AS REPORT |   |  |

UNCLASSIFIED

|  |   |   |
|--|---|---|
| <p>21a. NAME OF RESPONSIBLE INDIVIDUAL</p> <p>N. A. Hupp</p> | <p>21b. TELEPHONE (include Area Code)</p> <p>(619) 553-3655</p> | <p>21c. OFFICE SYMBOL</p> <p>Code 441</p> |
|  |   |   |

# INITIAL DISTRIBUTION

|           |                |      |
|-----------|----------------|------|
| Code 0012 | Patent Counsel | (1)  |
| Code 0141 | A. Gordon      | (1)  |
| Code 0144 | R. November    | (1)  |
| Code 44   | J. Grossman    | (1)  |
| Code 441  | C. Dean        | (6)  |
| Code 441  | S. Bemis       | (1)  |
| Code 441  | N. Hupp        | (12) |
| Code 952B | J. Puleo       | (1)  |
| Code 961  | Archive/Stock  | (6)  |
| Code 964B | Library        | (3)  |

Defense Technical Information Center  
Alexandria, VA 22302-0268 (4)

NOSC Liaison Office  
Washington, DC 20363-5100

Center for Naval Analyses  
Alexandria, VA 22302-0268

Navy Acquisition, Research & Development  
Information Center (NARDIC)  
Alexandria, VA 22333

National Security Agency  
Fort Meade, MD 20755

University of California, San Diego  
San Diego, CA 92186

Ameritech Services  
Rolling Meadows, IL 60008

Speech Systems, Inc.  
Tarzana, CA 91356